CROATICA CHEMICA ACTA CCACAA **64** (3) 371–388 (1991)

ISSN 0011-1643
UDC 541
CCA-2008

Original Scientific Paper

Regularities in Non-crystalline Hydrogen Bonded Systems: Water, Solutions and Biological Macromolecules 30 Years on

John Finney

ISIS Science Division, Rutherford Appleton Laboratory, Didcot, Oxon, OX11 0QX. UK

and

Hugh F. J. Savage

Chemistry Department, University of York, Heslington, York, YO1 5DD, UK

Received May 20, 1991

Major conceptual advances were made by Pauling several decades ago in using information from crystal structures to try to rationalise the structures of apparently complex hydrogen bonded systems. Thus were born not only the α and β structures now recognised as secondary structures in globular proteins, but also reference structures such as the clathrate cage which he used to try to understand the structure of liquid water. Recent developments are discussed which, following the same conceptual approach, have thrown new light on our understanding, not only of the stability of globular proteins, but also of the orientational structure of water at important biomolecular interfaces. The power of the structural regularities uncovered shows that the conceptual approach taken by Pauling in the 1950s retains its power today, and is far from being overshadowed by the impressive numerical work of computational quantum chemists and computer simulators.

INTRODUCTION: REGULARITIES IN DISORDERED STRUCTURES

Disordered hydrogen bond systems have been of strong interest for several decades. Liquid water is important not only as a chemical solvent, but also as the environment in which many biomolecular processes take place: its role is generally thought to be central to biomolecular stability and interactions. Biomacromolecules themselves are at first sight »disordered«, a globular protein consisting of at least one chain of amino acids which folds on itself to form an intricate, apparently complicated, set of (charged, hydrogen bonded, van der Waals, disulphide) interactions in its active, native structure.

The absence of a crystal lattice in such non-crystalline systems has meant that describing their structures – an essential prerequisite for understanding their properties and interactions – has not been straightforward. Early approaches to simple liquids attempted to characterise their structures with references to an underlying crystal lattice that was easy to describe in terms of the contents of a unit cell and a set of lattice translation operations. Such models are now recognised as oversimplifications, seriously underestimating the inherent irregularity and disorder characteristic of liquids in general. However, consistent with this inherent structural irregularity, liquids *do* have structures, albeit perhaps characterisable only statistically in terms of instantaneous structures, or average structural characteristics. As far as simple liquids of essentially spherical molecules are concerned, the nature of the characteristic structural regularities were graphically demonstrated by Bernal in the late 1950s[1] in his random close packed model of simple liquid structure. He demonstrated that it was possible to pack together equal spheres (representing spherical atoms) in a homogeneous, irregular arrangement containing no long range crystalline regions (see Figure 1). Examining such models in detail allowed him to begin to characterise the local structural regularities that dominated the liquid structure at short range.These were in fact simple polyhedra

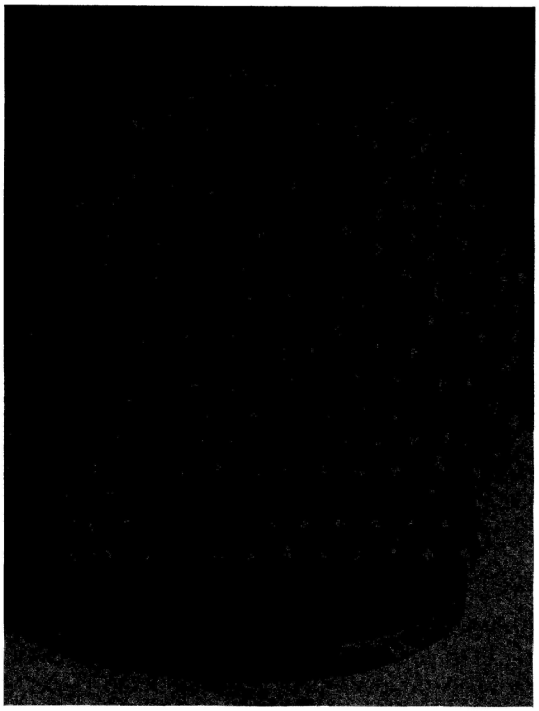

Figure 1. The intrinsic irregularity characteristic of a simple liquid (upper part) contrasted with the regular repeating structure of the corresponding crystal. In this simple liquid model of Bernal,[1] the atoms are represented by hard spheres.

such as tetrahedra and octahedra, local structural units which could be related to the geometrical characteristics of the interatomic interactions – in this case, the *spherical symmetry* of the repulsive core of *e.g.* the inert gas atoms that form such so-called simple liquids.

Ultimately, the essential geometrical characteristics of the intermolecular interactions must be deduced from other studies before we can, using this conceptual approach to non-crystalline structures, begin to try to understand the short-range regularities in non-crystalline structures. Two main sources of such information might be stressed. In quantum mechanical calculations, interaction energies may be estimated (with a degree of precision depending strongly on the complexity of the calculation) between pairs of molecules whose mutual configuration is varied to map out an energy surface, and hence both distance and orientation variations of pair interaction energies may be obtained. Thus a model pair potential may be constructed whose essential geometrical characteristics might be used to try to understand the local configurations occurring in non-crystalline systems. Secondly, *crystal* structures can be examined in detail to identify local molecular arrangements that occur frequently; provided the removal of the repeating lattice environment remains not too large a perturbation, such local structures could – if treated with care – give guidance as to the local structural regularities that may occur in non-crystalline systems.

Using such approaches, the foundation of our understanding of non-crystalline hydrogen bonded systems was made many decades ago by two of the century's great structural scientists, Bernal and Pauling. Bernal, together with Fowler, laid the basis of our understanding of water in a historical paper in 1933.[2] Using information then available on the nature of the water-water hydrogen bonded interaction, especially its geometry, together with structural information known from the then recently-solved structure of ice[3], they referred the structure of water to local characteristic structural aggregates that were consistent with interaction geometries and other known properties (in particular density changes with temperature). Adopting a similar approach 25 years later, Pauling suggested[4,5] the characteristic structures were – for good reasons of thermodynamic stability – different. Rather than accepting the model of Bernal and Fowler that was based on the local structures of quartz and tridymite found in silicates, proposed instead characteristic structures for liquid water that were related to the clathrate cage arrangements that had been found in inert has hydrates (work in which Pauling himself was involved, solving with Marsh[6] the structure of the so-called type I clathrate structure of choline hydrate).

Moving away from water to macromolecules such as globular proteins, it was Pauling in collaboration with Corey[7,8] who made the major step forward in beginning to rationalise these complex molecules whose structures also appeared to depend significantly on the geometry of the hydrogen bond interaction. Leaning heavily on his knowledge of hydrogen bond geometries built up from crystallographic studies of a variety of relevant crystal structures, the α-helix was born as a structural motif to be expected in folded protein molecules. The recognition of this regularity was a major step forward in rationalising the structures of these apparently complex molecules, and it set in train attempts to understand proteins in terms of secondary structural units that are very active today.

Both of these fields – namely, the structure of water and the structure of proteins – have moved a long way in the intervening years. Computers have revolutionised our ability to calculate pair potential functions (though perhaps there is still a long way

to go in understanding for example non-pair-additive effects, whose importance have been increasingly recognised since the early relatively crude attempts to allow for them in the 1970s[9-12]). X-ray crystallography has solved several hundred protein structures, allowing proteins to be discussed in terms of packing of α-helical regions, β sheets and turns. From these crystallographic studies, these structural regularities can be geometrically specified in some cases very precisely indeed, allowing distortions from the idealised units to be quantified, together with the potential interaction of them with the solvent. Neutron diffraction from hydrogen bonded crystals has enabled hydrogen atoms to be accurately located, with consequences for our detailed understanding of the geometry of the hydrogen bonding interaction. Finally, X-ray and – more importantly – neutron scattering studies, allied to computer simulation calculations again made possible through dramatic increases in computer power over 20 years, have thrown considerable light on the actual structures of both water and aqueous solutions.

It is not our intention here, however, to review this progress. Rather, we wish to stand on the shoulders of Pauling to look again at the power of structural regularities and the understanding of non-crystalline hydrogen bonded systems, using the approach of referring to information available from crystal structural studies. Two examples will be given, namely, the structure of water (in the liquid, but with a stress on water at interfaces with other molecules) and globular proteins. We will try to demonstrate that powerful, highly reliable structural regularities can be found from crystal structure studies that allow us to understand much more deeply not only the structure of water, both in the bulk and at important biological interfaces, but also the stability of globular proteins. We believe the approaches we discuss echo that taken several decades ago by Pauling in trying to understand both systems. The quality of the information now available may be much higher than was available to him, but the conceptual approach he took then retains its power today, and it is far from being overshadowed by the impressive abilities of the computational quantum chemists and simulators.

REFERENCE STRUCTURES IN WATER: BERNAL AND PAULING

In a paper published in 1933, Bernal and Fowler[2] laid the groundwork of our understanding of water and aqueous solutions. Written after the hydrogen bond had been identified, and the structure of ice had been established,[3] they proposed that the increase in density observed on melting the open, tetrahedral ice I structure (which is topologically similar to the structure of the silicate tidymite) may relate to the breakdown of part of the tridymite-like structure into aggregates that were quartz-like. These latter clusters would be more dense than the tridymite-like aggregates, and the relative fractions of the two kinds of aggregates would vary with temperature to explain the density changes observed. Bernal underlined many times that this model should not be taken too literally: he argued that water structure should be considered as an irregularly and tetrahedrally bonded set of water molecules,[13] for which the tridymite-like and quartz-like clusters provided a convenient first order description of the nearest neighbour arrangements. It was never intended as a model to explain the arrangement at greater than three of four molecular neighbours: rather, it attempted to give a rough picture of the breakdown of the tetrahedral coordination in the liquid as evidenced by the volume decrease up to the temperature of maximum density at 4 °C.[14]

Around 25 years later, Pauling proposed a modification of the Bernal and Fowler approach.[4] Again, he emphasised the irregularity expected in a liquid when he stressed »no doubt the structure has much randomness, but it is likely to contain certain con-

figurations with high frequency«.[5] However, he rejected on stability grounds the presence of quartz-like aggregates, stating that such a structure cannot be stabilised with respect to that of tridymite.[4] This was, he argued, because a quartz-like structure would have strained hydrogen bonds, with the O–H...O angles bent to around 135°, compared to the essentially unstrained straight (180°) H-bonds in the tridymite-like structure of ice. There were no entropic factors he could identify to compensate for this bond straining, and hence rejected the presence of such quartz-like aggregates in liquid water.

Pauling proposed instead a reference structure based upon the pentagonal dodecahedral water networks that had been found in many gas hydrate structures (the structure of the 12 Å unit cell structure of the chlorine hydrate – the so-called clathrate structure I – having been solved by Pauling and Marsh[6]). He argued[5] that »It seems not unreasonable to discuss the structure of water in terms of methane hydrate«, and proceeded to propose replacing each cage-centring methane molecule with a water molecule making no hydrogen bonds to the cage framework waters. Such a model he argued was consistent with several known properties of water. The model had 85% of the number of hydrogen bonds in ice, the lost 15% being consistent with the enthalpy of fusion of ice. From the methane hydrate dimensions, the water density of 1.00 gcm^{-3} was reproduced, and a »reasonable account« could be given of the dielectric dispersion and the X-ray diffraction pattern. It is again important to stress that a rigid crystalline-like structure was *not* envisaged: merely these pentagonal dodecahedra could be considered as major local structures in a fluid arrangement. As he stated in his original presentation of the model,[4] it was not necessary to retain the orientation of the dodecahedral cages found in the corresponding gas hydrate. The dodecahedra might have the same orientations, or perhaps be arranged differently, for example in ways more akin to fcc or hcp packings. In restating the model a little later,[5] he postulated that the dodecahedra could be arranged mutually in a large number of ways: consequently, the highly random structures for liquid water might be based on aggregates of water so bonded. The model was not envisaged as one which is stable and rigid, but labile.

REFERENCE STRUCTURES REVISITED: NEAR NEIGHBOUR REPULSIVE RESTRAINTS

Nearly 60 years on from Bernal and Fowler, and over 30 years on from Pauling's interstitial dodecahedral model, we might argue that many of the problems they experienced in communicating their ideas through such reference structures may have stemmed from the apparent conceptual difficulties of reconciling structural aggregates from crystals with the inherent irregularity they both knew was central to liquid structures. Although they both stressed again and again that their models were *not* ordered over long range, reconciling the existence of such »closed ring structures« as the quartz- and tridymite-like aggregates of Bernal and the dodecahedral cages of Pauling was not straightforward. Knowing what we now know about the inherent irregularity of liquids, we may feel that the reference structures chosen were essentially too large; nevertheless, the conceptual jumps involved in their thinking were major ones, which not all of us have yet managed to follow.

The subsequent history of the Pauling model is interesting in several respects. First, it provided Frank and Quist[15] with a model around which to build an influential statistical mechanical theory of water, in which (initially) the water molecules were considered to exist in one of two »states«, namely framework and interstitial. The properties of the framework and interstitial waters could be deduced, and many of them

found to be physically reasonable. Furthermore, it allowed the development of related concepts of structuring of water around non-polar groups in aqueous solution, clathrate-like cages consequently working their way into the conventional wisdom of apolar group hydration. As a model of water itself, however, the particular reference structures have failed to survive, being found inconsistent with, in particular, structural data from much improved X-ray scattering measurements of water.[16]

Present work on water – and other liquid – structure has moved away completely from such reference structure approaches. The conceptual need to think in terms of disordered crystalline aggregates has been removed (a major step forward in this direction being made by Bernal in his work on simple liquids[1] – see Figure 1). Modern computers allow us, with the assumption of an intermolecular potential function, to simulate the inherent irregularity that is consistent with the assumed intermolecular potential and the external conditions of temperature and pressure, and to make comparisons with experiment through the radial distribution function. A case might however be made for retaining some of the characteristics of the reference structure approach, especially in relatively complex aqueous systems, for example water in restricted environments such as close to a macromolecular surface. It is here that recent work which has identified *regularities* in disordered water structures promises to improve our understanding of the geometry of water structures – and other hydrogen bonding groups – relevant to biomolecular interactions. The regularities themselves may not be identifiable aggregates of molecules as in the clathrate reference structure of the Pauling water model, but the conceptual approach to be described owes much to that adopted by Pauling. As again in Pauling's approach to water, it also relies heavily on information obtained from detailed crystal structure analysis of a large number of hydrate crystals of a varied range of complexity.

From the crystal hydrate structures solved in which hydrogen atoms have been located we can examine relationships between particular bond angles and bond distances. In a major review on hydrogen bonded structures published over a decade ago. Olovsson and Jönsson[17] produced a series of plots similar to that shown in Figure 2 in which hydrogen bond distances are plotted against hydrogen bond angles. Although there is a scatter, there is a case for drawing through the plotted points a »correlation line« to represent the »ideal geometry« which a hydrogen bond might be expected to approach, other constraints permitting. Thus, we might use these rather soft restraints implied by this apparent distance and angle correlation in, for example, the placing of water molecule hydrogens close to proteins where we may already have located the oxygen positions from X-ray diffraction.

If we look in more detail however, the situation is not straightforward. First, the scatter of these points is uncomfortably wide, being significantly greater than the errors in the experimental determinations. Moreover, some bent H-bonds are short, while some long ones are relatively straight, whereas the generalisation we would extract from such plots is in fact the opposite. Furthermore, particular crystal structures cause problems. An example is that of α-cyclodextrin[18] for which Figure 3 shows one particular water (number 1) in an orientation which is not as expected from the plot of Figure 2. Examination of this structure shows that this water is essentially free to rotate to straighten the hydrogen bond shown to allow it to come closer to the »ideal structure« represented by the correlation implied in Figure 2. If we use the tendency towards this apparent ideal as a restraint in refining the α-cyclodextrin structure, we would in fact get the wrong answer.

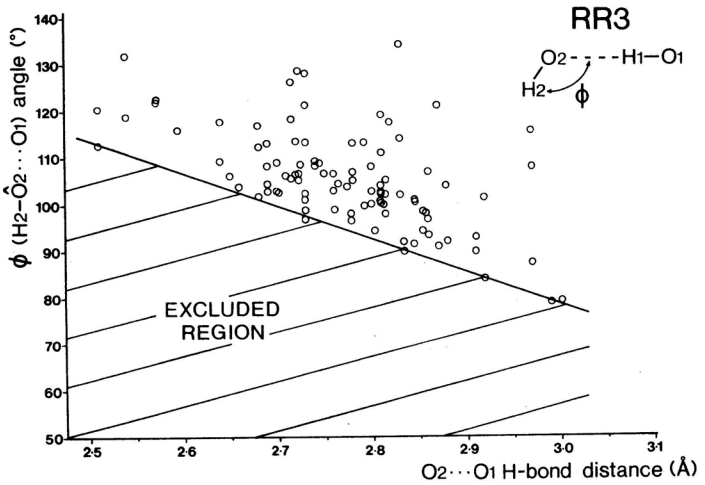

Figure 2. A plot of H2–O2...O1 hydrogen bond angles against O2...O1 H bond distances for water molecules involved in hydrogen bonds. A correlation line might be drawn through the points to represent an »ideal geometry« which hydrogen bonds might be expected to approach. The alternative approach presented here is the specification of the excluded region shown, in which hydrogen bond geometries are not found.

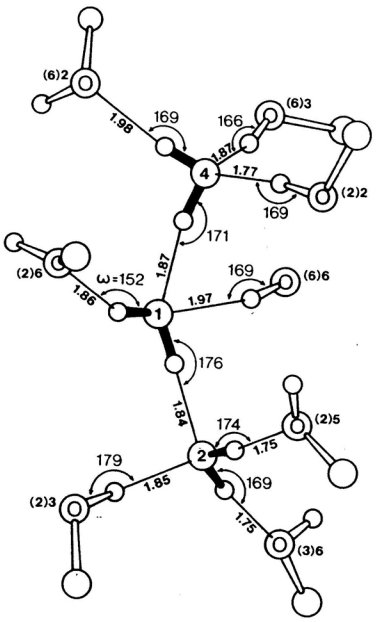

Figure 3. Hydrogen bond networks in a local region of α-cyclodextrin, showing the range of hydrogen bond distances (in Å), and in particular H-bond angles.

There thus appears to be something missing, and this missing element was uncovered through a high resolution X-ray and neutron diffraction study of crystals of vitamin B_{12} coenzyme, in which many of the hydrogen positions were clearly located.[19-21] Figure 4 shows part of the structure and one of the water networks that was located. Again, looking at the standard distances and angles shown in the figure, we find nothing new: they appear to be consistent with the scatter found in smaller structures and summarised in Figure 2. There is however another way to look at the water orientation in this system, and this is shown in Figure 5, where the same B_{12} region is shown but this time the distances between the oxygen centres and the non-hydrogen bonded hydrogen centres of the next neighbour water molecules are labelled. This may seem a very strange distance to identify, but doing so results in a very clear pattern indeed: the wide spread of distances and angles in Figure 4 is completely removed, and it seems that this oxygen-second neighbour hydrogen distance is constrained to be remarkably close to 3.0 Å.

Focusing on this apparently odd distance significantly modifies the approach we previously used to try to (not very successfully) rationalise water orientations. We have replaced a set of rather soft restraints on distances and angles which are of only limited use in helping us understand water orientations in this system by a very tight

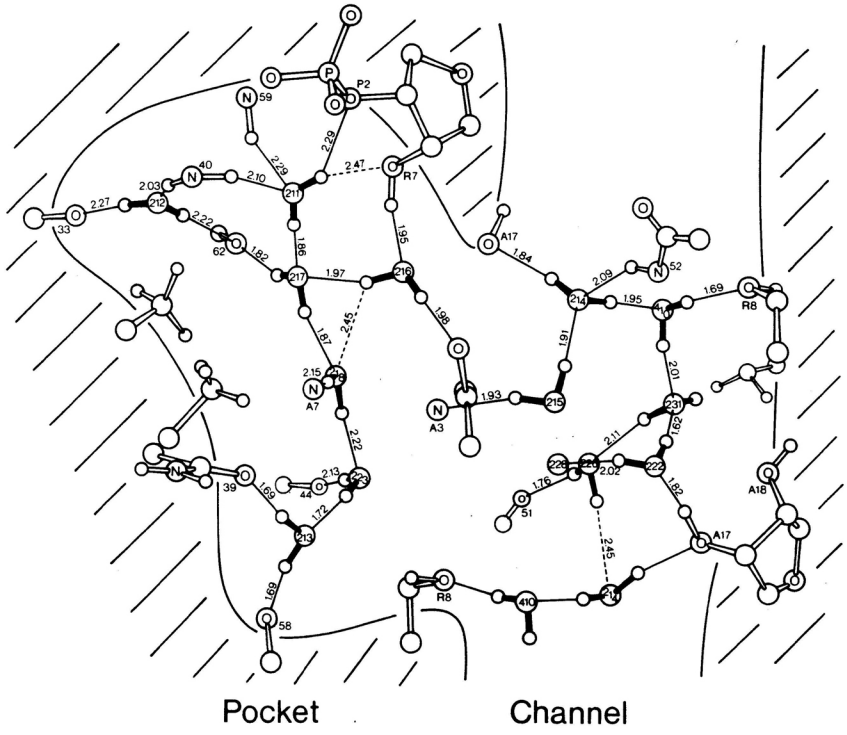

Figure 4. Main solvent networks in the pocket and channel regions of the coenzyme B_{12} hydrate. The H-bond distances specified show a wide scatter, consistent with the distance spread found from studies of smaller hydrates illustrated by the scatter of the points in Figure 2.

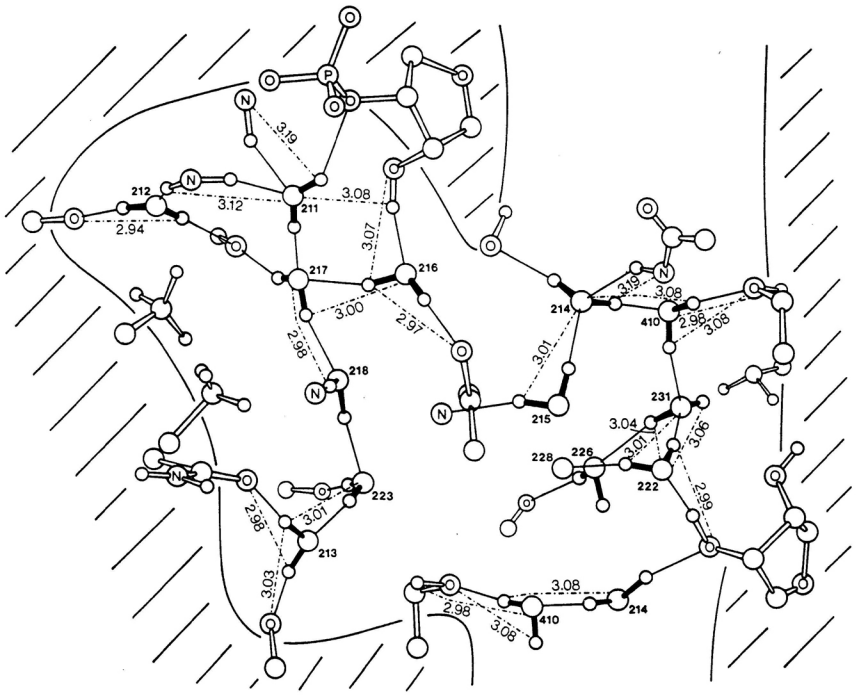

Figure 5. The same solvent network as shown in Figure 4, but this time with the close H2...O1 contacts given. Unlike the distances shown in Figure 4, the spread on these distances is small, suggesting a minimum oxygen-remote hydrogen distance of about 3.0 Å.

restraint indeed which – because of its tightness – can be much more powerful in helping us rationalise water structure. Looked at another way, we can replace the correlation implied in Figure 2 by an *excluded region* as shown. Now, instead of looking upon the implied distance-angle correlation as an ideal to which structures tend, we can assert that no configurations can exist which violate the excluded region. We have thus replaced a relatively weak correlation by a strong excluded (configurational) volume constraint which is inherently much more powerful.

The B_{12} example shown is not unique: since uncovering this *regularity* in water structure, a large number of hydrates and ice structures have been examined from the same viewpoint[22] and the results found to be wholly consistent with this minimum distance and the excluded volume region in Figure 2. Returning to the α-cyclodextrin case, Figure 6 shows the same region as Figure 3, though this time with the oxygen-second neighbour hydrogen distances labelled. The reason for the bent bond around the water 1 is now clear: to straighten it out would lead to the hydrogen from water 1 approaching O4 to less than the 3.0 Å.

Further investigation[20-22] shows that this *repulsive regularity* is one of four repulsive restraints (RR) (which are in fact related and can be unified in terms on a non-spherical van der Waals surface for the water molecule). These are illustrated in Figure 7 and discussed in more detail elsewhere.[20-23] Taken together, they define a set of ex-

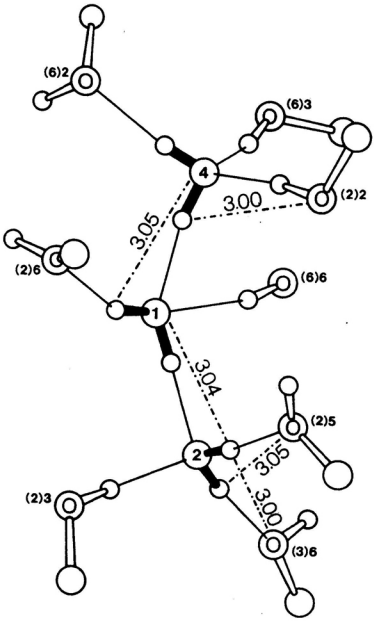

Figure 6. The cyclodextrin network of Figure 3, but with the minimum oxygen to remote hydrogen contacts shown.

cluded volume regions that allows us for the first time to rationalise the orientational structures found in hydrates in particular, and in hydrogen bonded systems in general. For example, the basic tetrahedral characteristics of water structure can be related to a minimisation of the RR repulsive interactions (especially RR3: remote H2...O1 contacts) and the maximisation of the number (4) and strengths (2.7–2.8 Å) of H bonds formed.[22] Shorter H bonds less than 2.7 Å require the acceptor water to adopt a more trigonal planar configuration to reduce the strain of the H2...O1 contacts, while longer

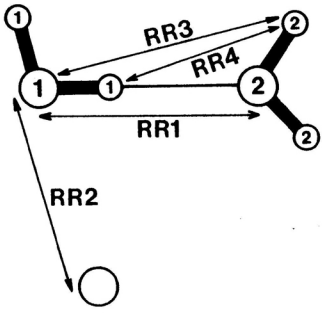

Figure 7. A schematic diagram showing the distances to which the four repulsive regularities (RRs) relate.

ones of 2.8–3.2 Å allow significant angular distortions from the tetrahedrality which is often required to maximise H bonding around the water molecules.

Whether the 3.0 Å minimum distance of RR3 is in fact the result of the van der Waals repulsion is open to discussion. For our use in understanding water both in the liquid and at complex interfaces such as B_{12} and proteins, it does not matter: the regularity itself is enough to allow us to be predictive in explaining local water structures, and even building up such structures from the incomplete data on hydration that in general comes from X-ray studies of protein structures. We might note in passing that the inclusion of what is in effect such an oxygen-remote hydrogen repulsive term in the MCY water potential[24] may perhaps be responsible for the relative success of this potential in reproducing the structure of water.[25] Moreover, in attempts to simulate the water in B_{12} coenzyme,[26] the water orientations are reproduced better when repulsive centres on the hydrogens are included in the potential function used.

This approach, which has resulted in very strong regularities in aqueous systems being established, has strong similarities to the approach taken by Pauling. From crystal structures, we have obtained information which we have used to try to generalise the structures of non-crystalline systems. Although we have not come up with reference structures as large as the clathrate cage in Pauling's water model, the regularities proposed are effective in helping us understand the non-crystalline structures. The scale of these regularities is shorter than those proposed by both Pauling and Bernal, and we do not build up relatively extensive topologically closed reference structures; rather we concentrate on the near neighbour interactions and make generalisations on local configurations that can be accessed in the aqueous structure. In that the data which facilitate this powerful rationalisation come from crystal structures, we are following an approach similar to that of Pauling. The approach is similar also in homing in on regularities that help us to describe, rationalise and perhaps even understand the disordered structures that we see. The scale of the regularities may be different – the RRs are restricted to near neighbour and next near neighbour correlations rather than to closed groups of atoms – but the conceptual approaches are similar and the result is a very useful set of powerful rules.

REGULARITIES IN PROTEINS

We now turn from the consideration of the labile structures of water to trying to understand regularities in globular proteins. As summarised above, this is an area in which Pauling and Corey made major advances in proposing the α-helix[7] and the β-sheet[8] structures. In doing this, they took notice of two main characteristics obtained by considering crystal structures of related systems, namely, the existence of the planar character of peptide bonds and the possibilities of hydrogen bonding between the main chain polar atoms. With respect to the former, Pauling's knowledge of resonance in chemical structures, from his work on the development of quantum chemistry from quantum mechanics in the 1920s, may have strongly influenced this appreciation of the planar/double bond character of the peptide bond that was an essential element in this pioneering work on secondary structures. It was a characteristic which Bragg, Perutz and Kendrew did not have in their published models[27] which all turned out to be incorrect. For the helix, Pauling also broke away from using an integral number of residues per turn, to which Bragg, together with most other groups pre-1950, adhered. In a way which is not unrelated, and which is similar to the approach we adopted above on repulsive regularities in water structure, we now look to see if we can

find further reliable consistent patterns that will allow us to rationalise the structure of not only polypeptide systems but also other relatively complicated biomolecular assemblies.

As in the case of Pauling, we return to crystallographic studies, but now to see what regularities we can observe in *protein* crystals. We find there are several. First, more than 90% of the polar atoms in the protein form hydrogen bonds either to the protein itself or – most importantly – to water. Thus there appears to be near maximisation of the hydrogen bonding capability, which seems to us to be a very strong tendency worthy of note. Secondly, very few polar side chains are buried in the interiors of proteins, and those that are form hydrogen bonds. We thus do not lose many hydrogen bonds through burying them inside the protein. Thirdly, protein cores are formed mainly of apolar groups, and fourthly, there are a significant number of apolar groups found on the surface.

Thus we appear to have a very strong constraint in protein structures in that the hydrogen bonding tends to be almost maximised. This is not surprising when we consider that the free energy of stability of a protein is equivalent energetically to the loss of only two to four hydrogen bonds. So even if there are compensating forces, the protein cannot afford to lose too many hydrogen bonds. This tendency towards near maximisation of hydrogen bonding appears stronger than the generally accepted need to bury apolar groups. In fact we have known for nearly 20 years[28] that in globular proteins up to about 50% of the apolar groups are *not* buried within the molecule, as would be expected from a strict application of Kauzmann's ideas of hydrophobic hydration driving protein folding.[29]

Concentrating on this tendency to make as many hydrogen bonds as possible, we now summarise what we call the lost hydrogen bonding (LHB) model, which helps us understand the complexity of protein structures in their folded state. To do this we first look at the stabilising factors which we believe are dominated by the tendency to make as many hydrogen bonds as possible, and to quantify this we look at several protein structures and evaluate the number of lost hydrogen bonds. *i.e.* those that would have been made in the extended state but are unmade in the folded protein. This quantitative assessment involves making comparisons between the native structure and the hypothetical extended chain, in which all the polar groups fully exposed to the solvent would make their full hydrogen bonded complements. The number of hydrogen bonds which are not made in the folded structure is a destabilising influence, and these are quantified for each of the protein structures considered.

Against this destabilising factor of lost hydrogen bonds, we set estimates of other »stabilising interactions«, namely, buried non-polar groups, made salt bridges and disulphide bonds. The latter two contributions are small but they do smooth the irregularities obtained if they are ignored. We now plot the destabilising lost hydrogen bond contributions against the stabilising factors above, and the result is shown in Figure 8. We interestingly observe a surprisingly strong correlation of 0.993 between these two factors, implying a compensation of the stabilising by the destabilising factors to a high degree of accuracy.

Similar to the excluded regions identified in hydrogen bonded structures in the previous section, this plot – or regularity – can now be used to help us to understand this particular kind of non-crystalline hydrogen bonded structure that is a protein. In particular, we can use this regularity to assess stability of protein structures, and as a structural check for correctness of non-covalent geometry within protein structures

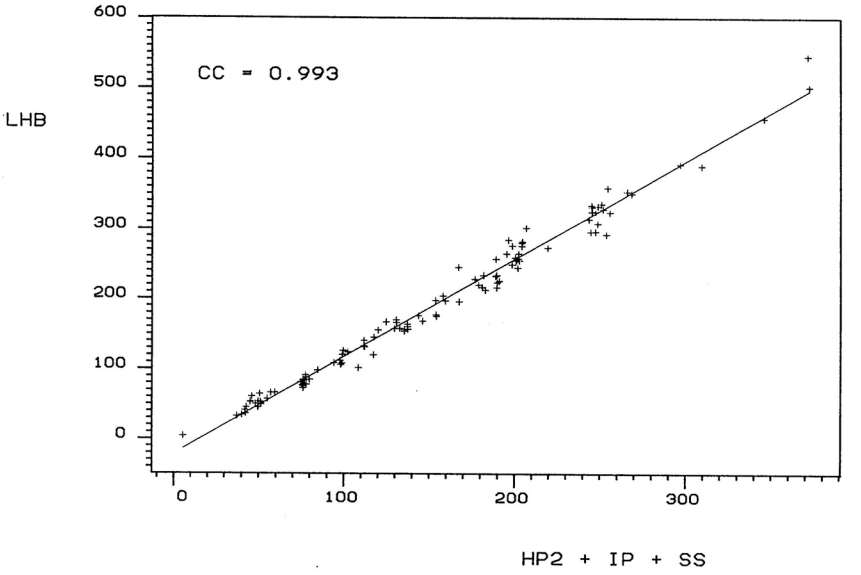

Figure 8. Plot of LHB (the sum of three lost hydrogen bond contributions) *versus* the sum of three stabilising interactions (HP: »hydrophobic energy«; IP: ion pairs; SS: disulphide bonds). The line represents a high correlation coefficient of 0.993.

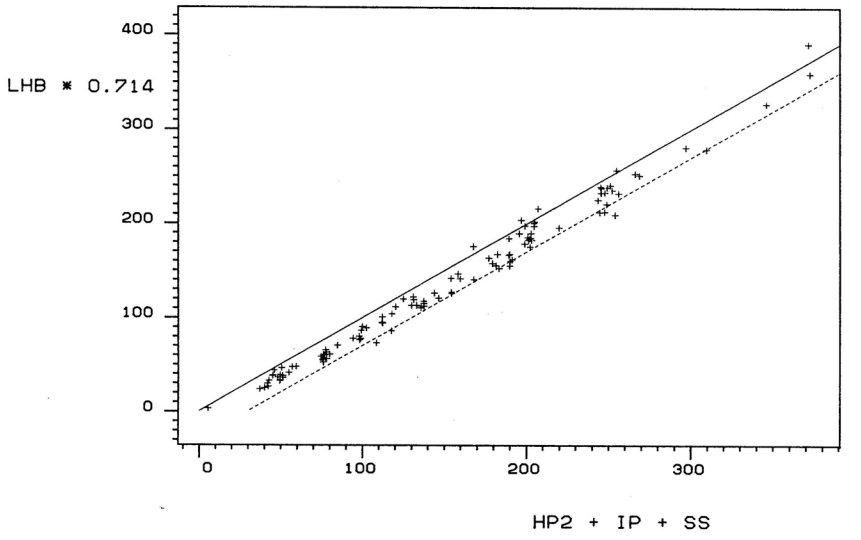

Figure 9. As Figure 8, but with the ordinate rescaled by 0.714. Most protein structures fall between the resulting line of unit slope and the dotted line limiting a window of 30 kcal mol^{-1}.

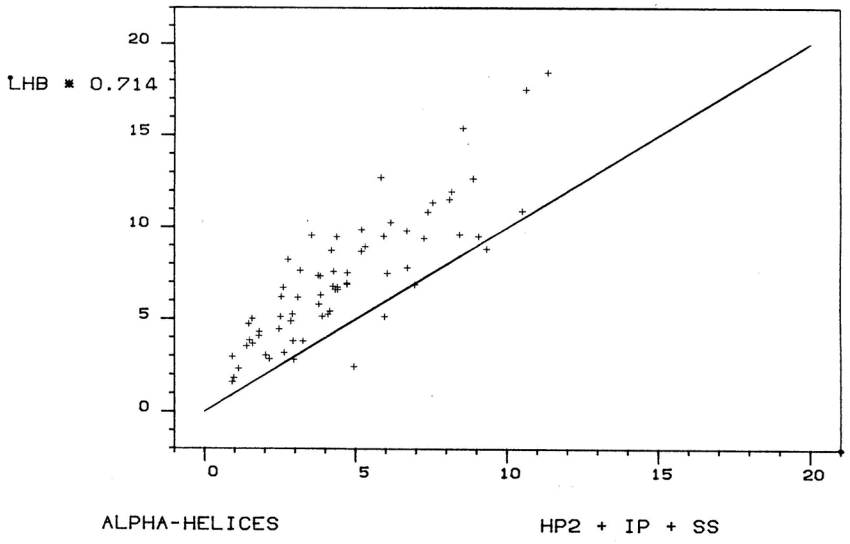

Figure 10. LHB plots for α-helical fragments from ~30 protein structures.

(particularly those known only at lower resolution). To facilitate this, we replot in Figure 9 the previous correlation plot, but additionally rescaling the vertical axis (LHB) by 0.714 such that the stabilising contributions just outbalance the destabilising ones. We find that the values for most proteins lie in a window of ~30 kcal mol^{-1} to the right of the unit slope line passing through the origin. If we now take a proposed structure, if its representative point on this plot falls below the correlation line, the lost hydrogen bond model predicts that the protein has a relatively stable structure. Where the point falls above the line, the proposed structure is unstable.

To illustrate the use of this regularity we give two examples. Figure 10 is the correlation plot for α helices isolated from about 30 native structures. Most points lie above the line, indicating that in isolation α helices are quite unstable, a result in agreement with solubility studies.[30] Of those helices that do appear below the line, one is the C-peptide of ribonuclease A, which is indeed reported to be experimentally stable.[31]

In contrast, a similar plot for isolated β sheets of between 2 and 6 strands from again about 30 protein structures is shown in Figure 11. In this plot most of the points lie below the line indicating such structures are fairly stable.

One particularly promising application of the above LHB model is as an effective filter procedure for assessing the stabilities of, for example, misfolded structures (perhaps in assessing theoretical folding procedures), and in estimating the likely stability or otherwise of mutations created, *i.e.* for biotechnological applications: there is after all no point in going to the effort of making a modified protein if it will not fold. The procedures explained here can allow us to assess the likelihood of its being able to fold in the same way as the unmodified protein. The approach might also be extended to include other solvents (such as those containing protein denaturants or protectants) to probe how the model protein structures may behave in these different environ-

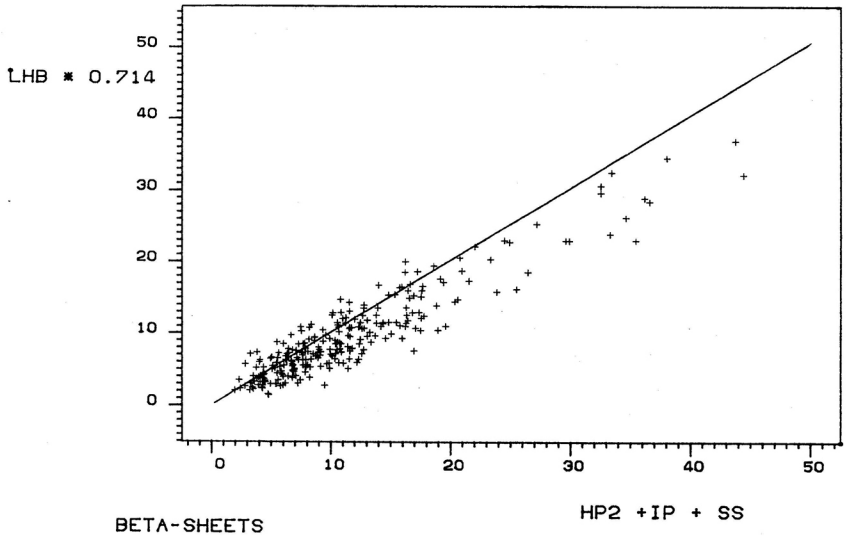

Figure 11. LHB plots for β-sheet fragments from ~30 protein structures.

ments, and possibly compare them with partly unfolded structures derived experimentally from 2D NMR.

This work on the lost hydrogen bond model – which is at a very early stage (though elements of it have been written up elsewhere[32]) – again builds on the approaches pioneered by Pauling to try to understand the structures and stabilities of proteins. As in his work on α helices and β sheets, we exploit information from crystallographic studies to pinpoint useful regularities such as the efficiency of hydrogen bonding and apolar group burial in globular proteins. Using these regularities, and including the all important solvent interaction possibilities, we are able to set up what appears to be a very powerful filter through which only stable proteins (or structural elements of proteins) can pass. Interestingly, we might note in passing that although the α helix was initially proposed by Pauling and Corey as a structural element, such units are apparently – often because of solvent effects – not stable when removed from the protein environment.

5. SUMMARY

This paper contains little that has not been published elsewhere. What it tries to do however is to put some recent developments into the framework of the approach pioneered by Pauling in using information from crystal structures to try to rationalise, and hopefully improve our understanding of, the complex systems that are aqueous liquids and globular proteins. With the major advances that have been made in radiation sources and instrumentation for X-ray and neutron diffraction studies, the database of local molecular configurations we have available to us is very much bigger than was available to Pauling in the 1950s. Moreover, those advances have enabled us to examine much more complex crystals (B_{12} is an example) which – because of effectively

reduced symmetry constraints locally – allow us to see water – and other hydrogen bonded molecules – in a very varied range of local configurations. Once the hydrogen atoms were located (an advance which requires neutrons) such work led to the identification of hydrogen bond repulsive regularities (RRs) which allow for the first time a consistent structural rationalisation to be made of hydrogen bonded structures at the molecular level. Insofar as we use information on local structures from crystals, the concept follows that of Pauling. However, it moves away from considering local »closed« reference structures (*e.g.* the clathrate structure of Pauling's interstitial water model, the quartz/tridymite units used in the model of Bernal and Fowler) to the »open« structures, or local geometrical configurations, that describe the more limited short range order. Understanding the structural restraints operating at this level seriously restricts the configurational space available to interacting hydrogen bonding molecules. As mentioned above, this replacement of a distance-angle correlation (Figure 2) with an excluded region of configurational space – taking advantage of excluded volume restrictions – is a powerful development promising much in the understanding of both water and macromolecule hydration, and hopefully also in clarifying the relevance of hydration to macromolecule stability and interactions.

Moving on to proteins, this theme can be developed further to rationalise protein stability. Building again in a similar way to the approach taken by Pauling for the α helix, and combining this with ideas developed from the repulsive regularity work on aqueous systems, strong regularities can be identified which allow the assessment of the stability of proteins and other molecular assemblies. In this approach, not only are the intramolecular hydrogen bonding possibilities that Pauling concentrated on in the 1950s important, but also the possibility of solvent interaction is brought in as a strong constraint, together with estimates of the effects on stability of disulphide bridges, ion pairing, and efficient packing of non-polar groups within the protein. Whether van der Waals forces or the classic hydrophobic interaction is relevant to this latter point is perhaps a matter of terminology, and is an argument we shall not raise further here. The LHB approach we develop is, we believe, potentially a major step forward with applications in biotechnology, for example in screening possible mutants for stability.

A POSTSCRIPT ON CLATHRATE STRUCTURES

Advances in neutron instrumentation and sources have recently allowed us to identify directly, we believe for the first time, the cage structures proposed by Pauling in his interstitial model of water. However, these have been observed not in water itself, where other experimental results preclude their existence, but in solutions of selected molecules. As mentioned in the main text, the conventional wisdom of solutions of apolar molecules implies there is some »ordering« of water around non-polar groups in solution, an ordering which is lost (with a consequent entropic gain to the system) when two such groups come together in a solution under the influence of the so-called hydrophobic interaction.

Previous attempts to directly identify such structures in appropriate solutions have been successful. Using neutrons – but from a pulsed spallation source rather than a fission reactor, for technical reasons – we have obtained information on the hydration of the tetramethylammonium (TMA) ion which is consistent with the existence of a disordered defective cage structure of water molecules. Thus, TMA, although being weakly charged, appears to hydrate as a classical non-polar molecule. More recent work

on methanol has also concluded that the exposed methyl groups are surrounded by (incomplete) cage structures, again with a geometry consistent with those found in clathrate hydrate crystals. Some of these results – which promise considerable extension relevant to biomolecular interactions as well as to fundamental solution chemistry – and the relevant technical details are presented elsewhere[33,34]

REFERENCES

1. J. D. Bernal, *Proc. Roy. Soc.***A280** (1964) 299.
2. J. D. Bernal and R. H. Fowler, *J. Chem. Phys.* **1** (1933) 515.
3. W. H. Barnes, *Proc. Roy. Soc.* **A125** (1929) 670.
4. L. Pauling, in D. Hadži (Ed.) *Hydrogen Bonding*, London, Pergamon (1959) p. 1.
5. L. Pauling, *The Nature of the Chemical Bond,* Oxford University Press, 3rd Edition (1960).
6. L. Pauling, and R. E. Marsh, *Proc. Nat. Acad. Sci. US.* **38** (1952) 112.
7. L. Pauling and R. B. Corey, *Proc. Nat. Acad. Sci. US.* **37** (1951) 235.
8. L. Pauling and R. B. Corey, *Proc. Nat. Acad. Soc. US.* **37** (1951) 51.
9. P. Barnes in H. J. C. Berendsen (Ed.) *Report of CECAM Workshop on Molecular Dynamics and Monte Carlo Calculations on Water,* CECAM, Orsay, France (1972) p. 77.
10. H. J. C. Berendsen and G. A. van der Velde, in H. J. C. Berendsen (Ed.) *Report of CECAM Workshop on Molecular Dynamics and Monte Carlo Calculations on Water,* CECAM, Orsay, France (1972) p. 63.
11. P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature* **282** (1979) 459.
12. J. L. Finney and J. M. Goodfellow, in E. Clementi and R. H. Sarma (Eds) *Structure and Dynamics: Nucleic Acids and Proteins, Adenine,* New York (1982) p. 81.
13. J. D. Bernal in *Proceedings of the First International Symposium on Water Desalination.* US Department of Interior (1965) p. 371.
14. J. D. Bernal in D. Hadži (Ed.) *Hydrogen Bonding,* London, Pergamon (1959).
15. H. S. Frank and A. S. Quist, *J. Chem. Phys.* **34** (1961) 604.
16. A. H. Narten and H. A. Levy, *Science* **165** (1969) 447.
17. I. Olovsson and P. G. Jönsson, in P. Schuster, H. Zundel, and C. Sandorfy (Eds) *The Hydrogen Bond,* Amsterdam, North Holland (1976) p. 393.
18. B. Klar, B. Hingerty, and W. Saenger, *Acta Cryst.* **B36** (1980) 1154.
19. F. J. Savage, *Biophys. J.* **50** (1986) 947.
20. H. F. J. Savage, *Biophys. J.* **50** (1986) 967.
21. H. F. J. Savage and J. L. Finney, *Nature* **322** (1986) 717.
22. H. F. J. Savage, in F. Franks (Ed.) *Water Science Reviews,* Vol. 2, Cambridge, Cambridge University Press (1986) p. 67.
23. J. L. Finney and H. F. J. Savage, *J. Mol. Struct.* **177** (1988) 23.
24. O. Matsuoka, E. Clementi and M. Yoshimine, *J. Chem. Phys.* **64** (1976) 1351.
25. J. L. Finney, J. E. Quinn, and J. O. Baum, in F. Franks (Ed.) *Water Science Reviews,* Vol. 1, Cambridge, Cambridge University Press (1985) p. 93.
26. F. Vovelle, J. M. Goodfellow, H. F. J. Savage, P. Barnes, and J. L. Finney, *Eur. Biophys. J.* **11** (1985) 225.
27. W. L. Bragg, J. C. Kendrew, and M. F. Perutz, *Proc. Roy. Soc.* **A203** (1950) 321.
28. B. Lee and F. M. Richards, *J. Mol. Biol.* **55** (1971) 379.
29. W. Kauzmann, *Adv. Prot. Chem.* **14** (1959) 1.
30. C.-S. C. Wu and J. T. Yang, *Biopolymers 27 (1988) 423.*
31. A. Bierzynski, P. S. Kim, and R. L. Baldwin, *Proc. Nat. Acad. Sci. US. 79 (1982) 2470.*
32. H. F. J. Savage and J. L. Finney, in A. Gomez Puyou (Ed.) *Biomolecules in Organic Solvents,* CRC Press (in press) (1991).
33. J. Turner, A. K. Soper, and J. L. Finney, *Mol. Phys.* **70** (1990) 679.
34. J. Turner, J. L. Finney, and A. K. Soper, *Z. Naturf.* **46a** (1990) 73.

SAŽETAK

Regularnosti u ne-kristalnim sustavima s vodikovom vezom: voda, otopine i biološke makromolekule nakon 30 godina

J. L. Finney i H. F. J. Savage

Novi strukturni podaci o globularnim proteinima, vodi i otopinama razmatrani su u svijetlu ideja o vodikovoj vezi koje je razvio L. Pauling prije nekoliko desetljeća. Otkrića novih strukturnih svojstava pokazuju da Paulingov koncepcijski pristup i dan-danas zadržava svoju vrijednost.